Design as a Fusion Problem

John E. Gray and A. Sunshine Smith-Carroll
Code Q-23 Electromagnetic and Sensor Systems Department
Naval Surface Warfare Center Dahlgren
18444 FRONTAGE ROAD SUITE 328
DAHLGREN VA 22448-5161

Rabinder N. Madan
Mathematical, Computer and Information Sciences
Code 311
Office of Naval Research
875 North Randolph Street
Arlington, VA 22203-1995

***Abstract*— Statistical Mechanics has proven to be a useful model for drawing inferences about the collective behavior of individual objects that interact according to a known force law (which for a more general usage is referred to as interacting units.). Collective behavior is determined not by computing $F = ma$ for each interacting unit because the problem is mathematically intractable. Instead, one computes the partition function for the collection of interacting units and predicts statistical behavior from the partition function. Statistical mechanics was unified with Bayesian inference by Jaynes who demonstrated that the partition function assignment of probabilities via the interaction Hamiltonian is the solution to a Bayesian assignment of probabilities based on the maximum entropy method with known means and standard deviations. Once this technique has been applied to a variety of problems and obtained a solution, one can, of course, solve the inverse problem to determine what interaction model gives rise to a given probability assignment. Probabilistic networks are important modeling tools in a variety of applications including social networks. We explore the usage of statistical mechanics as a mechanism to solve the inverse problem to determine the underlying interaction model that gives rise to the probabilistic network**

CONTENTS

keywords: Maximum Entropy; Design, Fusion; Connection Model

I. INTRODUCTION

Most academic discussions of distributed sensor networks focus on the problems associated with the fusion of data or the fusion of information. While these issues are important, they are not the only real issue along the path that motivates the reasons for distribution of sensors. Networking of the sensors enables the sharing of data from similar as well as disparate sensors, the fusing of the sensor data into information, and finally the synthesis of the information into knowledge. Information is power is a useful maxim only if it leads to better decisions which in turn result in better actions by the decisions maker. Sensor networks are creating tsunami of data that overwhelm but do not inform the end user of networks. With that in mind, we need to ask what are the methodologies that can dampen out the tsunami of data to achieve effective knowledge that leads to viable actions by the end user? While issues such as timeliness and ordering of the data, fusion methodologies for the data, reliability of the network, information flow within the network, delineating the difference between data and usable information are important; knowledge synthesis from the information provided by the network is the ultimate goal.

Distributed sensor networks are a topic of interest to both the military and civilians because they offer the possibility of increasing domain awareness in a given region. Complete situational awareness can occur only when sensors that operate at different frequencies and with different operational characteristics are effectively combined to provide a near-real time understanding of what is happening within a region. The synthesized "picture" can then be combined with information from other regions/domains to synthesize information that is of the type that is actionable– accurate and correct decisions can be formed based on the knowledge that has been synthesized from the distributed sensor networks. From the perspective of knowledge, however, the path to the future for networks that lead to knowledge synthesis is incomplete without the consideration of the semantic aspect of information— taking individual observations and judgments and making them part of the fusion problem. Ultimately, this fusion problem has to be considered from a perspective of an end user... someone whose career and life is on the line based on a decision to use or not use the information for certain ends.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JUL 2008		**00-00-2008 to 00-00-2008**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Design as a Fusion Problem	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Surface Warfare Center Dahlgren,Code Q-23 Electromagnetic and Sensor Systems Department,18444 Frontage Road Suite 328,Dahlgren,VA,22448-5161**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
11th International Conference on Information Fusion, June 30 ? July 3, 2008, Cologne, Germany.

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**8**	

From this perspective, decisions are more like designs, weaved from a tapestry of colors and patterns which are contradictory. When we think of a decision like a picture, the goals of the fusion problem become quite different, and almost unrecognizable. So how do we design a fusion problem? We don't know how to codify Design as a specific law, but we recognize it when we see it. A design is created from disparate elements that are combined to form a composition (true of artistic or engineering design). Each discipline has "atoms of composition" or atomic units that are fused together to form a specific instance of a design. In artistic designs, the method of fusion is based on the individual interpretations of the group aesthetic that falls under the aegis of "artistic temperament". For engineers, the group aesthetic must have some familiarity with the requirements of "physical reality". It is difficult to arrive at a mathematical criteria that helps clarify the direct problem of the process of composition.

"Probability is the logic of reasoning in the presence of uncertainty." Assignment of probabilities based on available information is Bayesian reasoning. This provides a solution of assignment of probabilities as a Maximum Entropy Problem (MEP). Thus a decomposition of a design into probabilities can be viewed as a solution to a MEP. Given that MEP is the method that provides the answer, then we can ask the question: "What is the Maximum Entropy problem that an assignment of design weights of the composition an answer to?" Being able to answer this question for "design classes" provides us a means to classify Design as a mathematical problem. Design algorithms can then be ordered in terms of the complexity of the problem solved. By providing the mathematical basis for synthesizing design as a composition, we hope to provide a pathway to analyze decisions from the perspective of design. This would allow decisions to be understood on a more quantative basis and ultimately bring the decision maker in the fusion process in a more analytical process.

II. Probability Issues Arising in Connection Models

A. Maximum Entropy Procedure

The *expected value* of the "surprise value" of a probability vector $|p\rangle$ is

$$H(n) = -\langle p|\ln p\rangle = -\sum_{i=1}^{n} p_i \ln p_i. \qquad \text{(surprise value)}$$

(See Papoulis for the details of the background in this section [9].) Note that for a deterministic probability vector, the entropy is zero, while for a uniform vector, the entropy is $\ln(n)$. The *degree of uncertainty* (which is equivalent to the surprise value) in the information is defined as

$$S[|p\rangle; n] = -k\langle p|\ln p\rangle = -k\sum_{i=0}^{n} p_i \ln p_i; \qquad \text{(Entropy)}$$

Most applications do not permit one to measure the probabilities p_i associated with a physical variable $f(x_i)$, instead the expected value $\langle f(x)\rangle$ is measured. Probabilities are connected to expected values by noting the formula

$$E[f(x)] = \langle p|f(x)\rangle = \sum_{i=0}^{n} p_i f(x_i) \qquad \text{(Expected Value)}$$

Maximum Entropy Method can be sumarized as: The goal is to find an assignment to the probability vector $|p\rangle$ subject to available information. Problem is to find an assignment of probabilities that maximizes the entropy. For example, we always know $|p\rangle$ is normalized

$$\langle 1|p\rangle = \sum_{i=0}^{n} p_i = 1, \qquad \text{(Condition 1)}$$

There is a known assignment of the means, α_r ($r = 1, 2, 3...m$) which satisfy:

$$\langle p|g_r(|x\rangle)\rangle = E[g_r(|x\rangle)] = \sum_{i=1}^{n} p_i g_r(x_i) = \alpha_r. \qquad \text{(Condition 2)}$$

The Method of Lagrange Undetermined Multipliers is a standard method for solving an optimization problem subject to constraints. The function we want to maximize is the entropy, so it is adjoined to the free parameters (Lagrange undetermined multipliers) times the equations for the constraints to form a **Lagrangian**. These two condition the principle "*The distribution, $|p\rangle$, that maximizes the uncertainty in the expected value subject to the constraint of the available information*" provides such an assignment that gives the probability assignment.

Example (Assignment of probabilities with no known information): The Lagrangian for entropy maximization is

$$L = -\langle p|\ln p\rangle - \lambda_0(\langle 1|p\rangle - 1)$$

Thus the least informative assignment of the probabilities when only the normalization condition is known is the assignment that the probabilities are uniformly distributed:

$$p_i = \langle p|e_i\rangle = \frac{1}{n}. \qquad (1)$$

This provides an explanation for **Occam's Razor**: "All things being equal, the simplest explanation is the best", e.g. uniform likelihood of assignment of truth to all explanations when wer are in a state of complete ignorance. **Einstein's Razor**: "Keep it as simple as possible–but not simpler." This implies that additional knowledge beyond the means is necessary to determine the probability assignment for propositions.

Example: The Lagrangian for entropy maximization (take $k = 1$) with known mean, α_r, is given by

$$\begin{aligned} L &= -\langle p|\ln p\rangle - \lambda_0(\langle 1|p\rangle - 1) \\ &\quad -\sum_{r=1}^{m} \lambda_r[\langle p|g_r(|x\rangle)\rangle - \alpha_r] \end{aligned} \qquad (2)$$

1) To minimize with respect to the probabilities, we compute the particular state $\langle p|e_i\rangle$ such that $\frac{\partial L}{\partial\langle p|e_i\rangle}$ is a minimum:

$$\frac{\partial L}{\partial\langle p|e_i\rangle} = 0 = \ln p_i + 1 + \lambda_0 + \sum_{r=1}^{m} \lambda_r g_r(x_i), \qquad (3)$$

2) Solving for the state's probability, $\langle p|e_i\rangle = p_i$, gives

$$\begin{aligned}\langle p|e_i\rangle &= e^{-\left(\lambda_0+1+\sum_{r=1}^{m}\lambda_r g_r(x_i)\right)} \\ &= e^{[-(\lambda_0+1+\langle\lambda|g(x_i)\rangle)]},\end{aligned}$$

(probability assignment)

3) Substituting probabilities into Condition 1 gives first Lagrange undetermined multiplier as

$$\begin{aligned}e^{(\lambda_0+1)} &= \sum_{i=1}^{n} e^{\left(-\sum_{r=1}^{m}\lambda_r g_r(x_i)\right)} \\ &= \sum_{i=1}^{n} e^{-\langle\lambda|g(x_i)\rangle}. \qquad (4)\end{aligned}$$

4) Other Lagrange undetermined multipliers are obtained by substituting the probability assignment into Condition 2

$$\langle p|g_r\left(|x\rangle\right)\rangle = \sum_{i=1}^{n} e^{-\langle\lambda|g(x_i)\rangle} g_r(x_i) = \alpha_r e^{(\lambda_0+1)} \qquad (5)$$

which solves the problem of assigning the probabilities.

5) By defining the partition function as Z

$$e^{(\lambda_0+1)} = Z, \qquad \text{(partition function)}$$

6) the Hamiltonian H is defined as

$$\begin{aligned}H_i &= \sum_{r=1}^{m}\lambda_r g_r\left(x_i\right), \qquad (6)\\ &= \langle\lambda|g(x_i)\rangle. \qquad (7)\end{aligned}$$

7) The probability assignment becomes

$$\langle p|e_i\rangle = p_i = \frac{\exp(-H_i)}{Z}, \qquad (8)$$

which is a familiar expression to physicists.

Thus, knowledge of the variance leads to be an exponential assignment to the probabilities. The physical connection suggested this solution allows one to connect probabilty assignments to physical interaction models as we now show.

B. Maximum Entropy Applied to Connection Models

A connection model is a technique to that allows us to breakdown or create an infrastructure in terms of information sources/sinks and links which may or may not "share" the information. Thus we tend to think of the mathematical structure in graph theoretical terms. An alternative way to think about this is in physical terms where the links represent interactions, while the nodes are sources for interactions such as a particle "interacting" with another particle.

Note, these notions are adopted from Harary [4]. There exists a **connection from the source (particle)** i **to** j if there is a sequence of sources $i,\ i_1, i_2, ..., i_k,\ j$ such that $(i,\ i_1), (i_1, i_2), ..., (i_k, j)$ that have some type of connection. If there are connections between each of the $\binom{n}{2}$ distinct pairs of nodes, then the connection model is said to be **totally connected**. An **adjacency matrix ,** A, for an $m-$edge connection model is the matrix

$$A = \begin{array}{c} \begin{matrix} a_{11} & a_{12} & \cdots & a_{1m-1} & a_{1m} \end{matrix} \\ \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & \cdot & 1 \\ 1 & 1 & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot \\ 0 & 1 & \cdot & \cdot & 0 \end{bmatrix} \\ \begin{matrix} a_{m1} & a_{m2} & \cdots & a_{mm-1} & a_{mm} \end{matrix} \end{array} \qquad (9)$$

where $a_{ij} = 1$ if source i is connected to sourse j and zero if it isn't. A source is not connected to itself (no self-interaction), so $a_{ii} = 0$. An adjacency matrix is *symmetric*, so $A = A^t$.

A random connection model can be defined in a variety of fashions with the randomness used to condition some aspect of the connection model. Assume a fixed set of vertices with a rule for assigning connection to the sources. A connection model with a source set $G = \{1, 2, ..., n\}$ and edge connection state $E = \{i : X(i), i = 1, 2..., n\}$ where $X(i)$ are independent random variables with a probability rule for assignment of a connection for $(j = 1, 2, ..., n)$:

$$P\{X(i) = j\} = f(j) \qquad (10)$$

The simplest rule for a connection is $f(j) = const.$, which is a uniform distribution or equal probability assignment of a connection between sources. For each particle i, randomly select a connection j and randomly decide to connect a particle to another particle from the source set. A connection model constructed in such a manner is referred to as a **random connection model**.

Example: For an arbitrary connection model that has a weight, w_{ij}, associated with each connection between source (i, j), we can associate a number that can be defined as a probability by the assignment:

$$P_{ij} = \begin{cases} \frac{w_{ij}}{\sum_j w_{ij}} & \text{if } (i,j) \text{ is connected} \\ 0 & \text{if } (i,j) \text{ is not connected.} \end{cases} \qquad (11)$$

Definition: One can define the **Gibbs entropy** of an ensemble, $\mathcal{E}$, of connection models as

$$S^{\mathcal{E}} = -\sum_{G\in\mathcal{E}} P(G)\ln(P(G)), \qquad (12)$$

subject to the normalization requirement that the probability of instances of the connection model that occur in an ensemble $\mathcal{E}$ as

$$\sum_{G\in\mathcal{E}} P(G) = 1. \qquad (13)$$

Example: Using MEP formalism with known prior information about the particular property of a connection model, $x_r(G)$, gives

$$\ln(P(G)) + 1 + \lambda_0 + \sum_{G\in\mathcal{E}}\lambda_i x_r(G) = 0. \qquad (14)$$

1) The **probability assignment for the connection model** is

$$P(G) = \frac{\exp(-H(G))}{Z(G)}. \qquad (15)$$

2) The **Hamiltonian of a connection model** is

$$H(G) = \sum_{i=1}^{n} \lambda_i x_i(G). \tag{16}$$

3) The **Partition Function of the connection model** is

$$Z(G) = \exp(1 + \lambda_0) = \sum_{G \in \mathcal{E}} \exp(-H(G)). \tag{17}$$

4) The Interpretation of an instance of a connection model, G, is a sample drawn from an ensemble $\mathcal{E}$ of connection models.
5) This definition is equivalent to that of a physical system defined over its microstate, provided the temperature is finite.
6) It is always possible to compute the expected value of any connection property of the ensemble, $x(G)$, over the particular model as

$$E[x(G)] = \sum_{G \in \mathcal{E}} P(G)x(G). \tag{18}$$

7) The way of arriving at a definition of the Hamiltonian means we have a way of specifying the interaction model for our network when we have a probability model for the connections instead of "physical model".
8) *We do not need to compute the statistical averages over the connection probability distributions directly, instead we can work from the definition of the partition function of our connection model, which is often easier.*

C. Some Observations Based Upon the Connection Model Partition Function

When one defines the partition function $Z(\mu)$, it is convient to introduce μ, a scaling parameter or the inverse temperature for a physical system. For a collection of data $\{f(x_i)\}$ with a probability vector, $|p\rangle$, the partition function is

$$Z(\mu) = \sum_i \exp(-\mu p_i f(x_i)) = \exp(-\mu \langle p|f(x)\rangle). \tag{19}$$

Take first derivative of $Z(\mu)$ with respect to μ

$$\begin{aligned} -\frac{\partial}{\partial \mu}[Z(\mu)] &= \langle p|f(x)\rangle \exp(-\mu \langle p|f(x)\rangle) \\ &= \langle p|f(x)\rangle Z(\mu). \end{aligned} \tag{20}$$

Rearrangement of this expression gives the definition of the expected value

$$-\frac{1}{Z(\mu)}\frac{\partial}{\partial \mu}[Z(\mu)] = \langle p|f(x)\rangle = \langle f\rangle, \tag{21}$$

which is the definition of mean of the data set $\{f(x_i)\}$. Thus, if one can find the partition function for a data set, the mean can be calculated from it.

A simple observation is that the left hand side of this equation is equivalent to

$$-\frac{1}{Z(\mu)}\frac{\partial Z(\mu)}{\partial \mu} = -\frac{d}{d\mu}(\ln Z). \tag{22}$$

The definition of the **free energy** is

$$F_\mu = -\ln(Z(\mu)), \tag{23}$$

so

$$\langle f\rangle = \frac{dF_\mu}{d\mu}. \tag{24}$$

Take the second derivative of F_μ, which gives

$$\begin{aligned} & -\frac{\partial}{\partial \mu}\left(-\frac{1}{Z(\mu)}\frac{\partial Z(\mu)}{\partial \mu}\right) \\ = & -\frac{\left(\frac{\partial Z(\mu)}{\partial \mu}\right)^2}{Z^2(\mu)} + \frac{1}{Z(\mu)}\frac{\partial^2 Z(\mu)}{\partial \mu^2} \\ = & -\langle f\rangle^2 + \langle f^2\rangle \\ = & \sigma_f^2. \end{aligned} \tag{25}$$

the expression for **variance** is the second derivative of the free energy:

$$\frac{d^2 F_\mu}{d\mu^2} = \langle f^2\rangle - \langle f\rangle^2. \tag{26}$$

The variance, σ_f^2, can be used to define the **heat capacity** of the network function f is

$$C_f = \mu^2 \sigma_f^2. \tag{27}$$

Note, there are a variety of physics based variables that can be defined as derivatives of the partition function including pressure, average energy, capacity, temperature, entropy, etc. These statistical mechanics variables may be worth considering in a broader context.

Any characteristic, C, of a connection model can find expected value from the formula

$$\langle C\rangle = Z^{-1}(G)\sum_G C(G)\exp(-H((G))). \tag{28}$$

To measure the parameter C, we simply adjoin it as a term added to the Hamiltonian

$$H(G) = \sum_{i=1}^{n} \mu_i \nu_i(G) + \theta C. \tag{29}$$

Calculating the expected value, $\langle C\rangle$, for a linear interaction potential from the free energy as

$$\langle C\rangle = \left.\frac{dF}{d\theta}\right|_{\theta=0}, \tag{30}$$

this generalizes to enable us to calculate correlation functions of the connection model as

$$\langle \nu_i \nu_j ... \nu_l\rangle = Z^{-1}\left[\frac{\partial}{\partial \mu_i}\frac{\partial}{\partial \mu_j}...\frac{\partial}{\partial \mu_l}\right] Z. \tag{31}$$

We can then calculate the correlation between two elements as

$$\frac{\partial^2 F}{\partial \mu_i \partial \mu_j} = \langle \nu_i \nu_j\rangle - \langle \nu_i\rangle \langle \nu_j\rangle, \tag{32}$$

while the extension to higher orders is straight forward. Thus, instead of interpreting $\lambda_i x_i(G)$ as the Lagrange undetermined multiplier times some property of the connection model, it can

be interpreted as an interaction potential between source, i, and source j. The λ_i can be interpreted as scaling parameter μ, a field coupling parameter, or the inverse temperature. It is possible to posit a variety of interaction models, work out their consequences and work out the equivalent probability distributions for them as well.

III. Synthesisizing An Interaction Model as a Solution

The more we know, the more complicated the form of the probability assignment becomes up to a point. In creating a network that fuses weighted information from different sources we are creating the equivalent of a physical interaction system that has an underlying physics that we can strive to understand and exploit this knowledge. Thus, an algorithm when looked at this way is a physics model based on interpreting our weighing of the data from different sources as an underlying interaction model. Any assignment of probabilities can always be viewed as the solution to a Bayesian Maximum Entropy problem. When this information is combined in a state vector, we have the possibility of hidden laws emerging. Given that is the case, then we can ask the question: "What is the Maximum Entropy problem that an assignment that a weighting scheme is the answer to?" For other examples not related to this question, see the suggestive paper by Kesavan [8].

A. Partition Functions. the Free Energy, and Models for $H(G)$

Graph theory has been used as a model for social networks since the eighties. The idea of applying statistical mechanics dates back to the seminal paper by Frank [3] and Strauss [14], while a recent overview is found in [13]. Park [10] has developed a detailed statiscal mechanics study of graph models. This work is adpoted from these three sources. We have introduced the notion of connection models rather than a graph model because the notion of a connection model is more evocative than a graph. Also, it is consitent with the older body of work reflected in Ross on the usage of probability models in the study of general probability models [11].

The simplest form for the Hamiltonian for a connection model is a model where we know the expected number of interactions for the model, namely: $\langle m \rangle$, so $H(G)$ becomes

$$H_\mu(G) = \mu m(G). \qquad \text{(Model 0)}$$

This model is trivial, so the next simplest Model is when adjacency matrix $A(G)$ has components a_{ij}

$$a_{ij} = \begin{cases} 1 & \text{if source } i \text{ is connected to source } j \\ 0 & \text{if source } i \text{ is not connected to source } j \end{cases}. \tag{33}$$

The number of interactions, m, is given by

$$m(G) = \sum_{i<j} a_{ij}(G), \tag{34}$$

while the Hamiltonian is

$$H(G) = \mu \sum_{i<j} a_{ij}(G). \qquad \text{(Hamiltonian Model 1)}$$

The partition function (PF) is

$$\begin{aligned} Z_\mu(G) &= \sum_G \exp(-H(G)) \\ &= \sum_{a_{ij}} \exp(-\mu \sum_{i<j} a_{ij}(G)) \\ &= \prod_{i<j} \sum_{a_{ij}=0}^{1} \exp(-\mu a_{ij}(G)) \\ &= \prod_{i<j} (1+\exp(-\mu)) \\ &= (1+\exp(-\mu))^{\binom{n}{2}}. \end{aligned} \tag{35}$$

The Model 1 free energy is given by

$$F_\mu^{(\text{Model 1})} = -\ln\left(Z_\mu^{(1)}(G)\right) = -\binom{n}{2}\ln(1+\exp(-\mu)), \tag{36}$$

while the mean is

$$\langle f \rangle^{(\text{Model 1})} = \frac{dF_\mu}{d\mu} = \frac{\binom{n}{2}}{(1+\exp(\mu))}. \tag{37}$$

The standard deviation is

$$\sigma_f^2(\text{Model 1}) = -\frac{\binom{n}{2}\exp(\mu)}{(1+\exp(\mu))^2}; \tag{38}$$

and the heat capacity is

$$C_f^{(\text{Model 1})} = -\mu^2 \frac{\binom{n}{2}\exp(\mu)}{(1+\exp(\mu))^2}. \tag{39}$$

To compare the interaction model to the underlying connection interaction probability model, we reparameterize p as

$$p = \frac{1}{(1+\exp(\mu))}, \tag{40}$$

so

$$\exp(\mu) = \left(\frac{1-p}{p}\right), \tag{41}$$

and therefore

$$p = \frac{\exp(-\mu)}{(1+\exp(-\mu))} \tag{42}$$

Thus, solving for $(1-p)$,

$$p\exp(\mu) = (1-p) = \frac{1}{(1+\exp(-\mu))}. \tag{43}$$

The first moment is

$$\langle f \rangle^{(1)} = \binom{n}{2} p, \tag{44}$$

while the standard deviation is

$$\sigma_f^2(1) = \binom{n}{2}(p-p^2). \tag{45}$$

The heat capacity is given by

$$C_f^{(1)} = \binom{n}{2}\left(\ln\left(\frac{1}{p}-1\right)\right)^2 (p-p^2). \tag{46}$$

The partition function is

$$\begin{aligned} Z_\mu(G) &= \exp\left(-F_\mu\right) = \left(1+\exp\left(-\mu\right)\right)^{\binom{n}{2}} \\ &= (1-p)^{\binom{n}{2}}. \end{aligned} \tag{47}$$

Therefore the probability, $P(G)$ for the connection model ensemble is

$$\begin{aligned} P(G) &= \frac{\exp\left(-H\right)}{Z} = \frac{\exp(-\mu m)}{\left(1+\exp\left(-\mu\right)\right)^{\binom{n}{2}}} \\ &= \left(\frac{1-p}{p}\right)^{-m}(1-p)^{\binom{n}{2}} \\ &= p^m(1-p)^{\binom{n}{2}-m}. \end{aligned} \tag{48}$$

This is the probability for a graph for which each of the $\binom{n}{2}$ possible edges appear with probability p, which is the well known Bernoulli random graph model studied by Erdos and Renyi.

B. General Model

A more general model is to replace μ_i with μ_{ij} so the Hamiltonian is

$$H(G) = \sum_{i<j} \mu_{ij} a_{ij}(G), \qquad \text{(General Model)}$$

then the partition function

$$Z_{\vec{\mu}}^{(\text{Model 3})}(G) = \prod_{i<j}\left(1+e^{-\mu_{ij}}\right). \tag{49}$$

The free energy is

$$\begin{aligned} F_\mu^{(\text{Model 3})} &= -\ln\left(Z_\mu^{(\text{Model 3})}(G)\right) \\ &= -\sum_{i<j}\ln\left(1+e^{-\mu_{ij}}\right); \end{aligned} \tag{50}$$

therefore the first moment is

$$\langle f\rangle^{(\text{Model 3})} = \frac{dF_\mu^{(\text{Model 3})}}{d\mu} = \frac{1}{\left(1+e^{\mu_{ij}}\right)}, \tag{51}$$

the standard deviation is

$$\sigma_f^2\ (\text{Model 3}) = \frac{d^2F_\mu^{(\text{Model 3})}}{d\mu^2} = -\frac{e^{\mu_{ij}}}{\left(1+e^{\mu_{ij}}\right)^2}, \tag{52}$$

and the heat capacity is

$$C_f^{(\text{Model 3})} = -\mu_{ij}^2\frac{e^{\left(\mu_{ij}\right)}}{\left(1+e^{\left(\mu_{ij}\right)}\right)^2}. \tag{53}$$

Then we can reparameterize p as

$$p_{ij} = \frac{1}{\left(1+e^{\mu_{ij}}\right)}. \tag{54}$$

Note, some simplified models beyond Model 1 are special cases with $\mu_{ij} = \left(\mu_i+\mu_j\right)$ for Model 2 and $\mu_{ij} = \mu$ for Model 1.

In general, one can specify interaction Hamiltonian as

$$H = \sum_{i<j}\Lambda_{ij}\sigma_{ij}, \qquad \text{(General Hamiltonian Model)}$$

where Λ_{ij} is parameter that couples each source together and is the connection sequence satisfies $\nu_i = \sum_j \sigma_{ij}$. The partition function is

$$Z = \prod_{i<j}\left(1+e^{-\Lambda_{ij}}\right), \tag{55}$$

while the free energy is

$$F = -\sum_{i<j}\ln\left(1+e^{-\Lambda_{ij}}\right). \tag{56}$$

Note the probability of an interaction between source i and j is the expected number of the connection sequence, then

$$p_{ij} = \langle\sigma_{ij}\rangle = \frac{\partial F}{\partial\Lambda_{ij}} = \frac{1}{\left(1+e^{\Lambda_{ij}}\right)}. \tag{57}$$

Underlying basis for the parameterization that gives us Bernoulli model with

$$P(G) = p_{ij}^m(1-p_{ij})^{\binom{n}{2}-m}. \tag{58}$$

Note, the term $\sum_i C_i\theta_i$ can be added to Hamiltonian without changing the probability model we have derived. This enables us to relate $\langle C_i\rangle$ to characteristic of the interaction model. Furthemore, the Bernoulli probability model derived from different types of interaction are all the same. So, one can conclude that *any interaction model that can be cast into the above form is a hidden Markov model.* This observation allows us to connect many seemingly disparate facts.

C. Example: Decision Fusion

A design is created from disparate elements, that when combined form a composition (true of artistic or engineering design). Each discipline has "atoms of composition" or atomic units that are fused together to form a specific instance of a design. In this instance, the rules for assigning connections are the rules of composition. In artistic designs, the method of fusion that constitute the finished product is based on the individual interpretations of the group aesthetic that falls under the aegis of "artistic temperament". For engineers, the group aesthetic must have some familiarity with the requirements of "physical reality". While it is difficult to arrive at a mathematical criteria that helps clarify the direct problem of the process of composition, the inverse problem of determining how the composition was generated is possible because we can infer law based on the probability assignment.

Our goal in this paper is to outline the mathematics of composition of decisions, which we have already done. Decision Composition can be viewed as a fusion of individual atoms of composition which are weighted components with these properties:

1) The numerical weights of the atoms of composition of a design can be assigned values between zero and one.
2) The sum of the weights can be normalized, so the sum of the weights of all the atoms of composition in an instance of a design add up to be one.
3) These two properties are mathematically equivalent to an assignment of probabilities, so they can be treated as probabilities.

For example, a particular design can always be viewed as an instance of, or a particular sample drawn from a space of possible compositions that have common elements. Proof of a mathematics theorem can also be viewed as a design with elements drawn from axiomatic systems combined with the glue of the rules of inference. Thus conceptually it is possible, in principle, to consider a proof as a composition relative to some system of music or as a painting. From the Bayesian perspective, any sufficiently complex problem has incomplete knowledge for us to try and reason through.

Reasoning from the perspective of *Lack of Knowledge* is equivalent to probabilistic reasoning. Laplace, Bayes, Gibbs, Shannon, Jaynes are source of this idea [5],[6], [7], [12]. It has proven to be the key to creating a real Information Age. "*Probability is the logic of reasoning in the presence of uncertainty.*" Assignment of probabilities based on available information is Bayesian reasoning. This provides a solution of assignment of probabilities which is a Maximum Entropy Problem (MEP). Thus a decomposition of a design into probabilities can be viewed as *a* (*there is no the*) solution to a MEP. Given that MEP is the method that provides the answer, then we can ask the question: "*What is the Maximum Entropy problem that an assignment of design weights is the composition an answer to*?" Being able to answer this question for "design classes" provides us a means to classify Design as a mathematical problem. Design algorithms can be then be ordered in terms of the complexity of the problem solved. By providing the mathematical basis for synthesizing design as a composition, we hope to provide a pathway to analyze decisions from the perspective of design. This allows decisions to be understood on a more quantities basis and ultimately bring the decision maker in the fusion process in a more quantitative process. This is illustrated with a simple example.

There are three ways to think about how decisions are synthesized: a branching process, a fusion problem from multiple information sources, or a grey scale picture where the lighting reflects the degree of connection between the individual components. Each of these three methods is useful depending on the problem domain.

The visual model is a way of reconstructing a complicated process where there are no particular analytical details known for its construction. By assigning different weighting schemes we can see a decision emerges out of a morass of disparate data.

For a branching process, the starting point can be thought of as the final decision, and the branches constitute the elements that went into the decision. At each of the branching levels the decisions have weights that must normalize like probabilities, so the formulation of the connection model adjacency matrix, A, reveals structure associated with the retrodiction process. The structure of the operator algebra gives an algebraic understanding of decisions in terms of "operator logic" of the matrix.

For fusion problems associated with target tracking, the third method is most useful. *The asynchronous multi-sensor track fusion problem can be stated as follows: Given a number of asynchronous valid tracks,* $\{X_i(t_{k_i}|t_{k_i}),P_i(t_{k_i}|t_{k_i})\}$, $i = 1, 2, ..., n$, *that arrive at the fusion center during the time interval* $[t_{k-1}, t_k]$, *find the best estimate in the minimum mean square sense of the system state at time,* t_k, *when it is computed according to the fusion rule.*

$$X_f(t_k|t_k) = \sum_{i=1}^{n} L_i(t_{k_i}) X_i(t_{k_i}|t_{k_i}) \tag{59}$$

where the L_i*'s are unknown weighting matrices to be determined.* The two constraints imposed on the fused track are unbiased if

$$\sum_{i=1}^{n} L_i \Phi(t_{k_i}, t_k) = I. \tag{60}$$

This represents the first constraint on the choice of the weighting matrices. If we solve for L_n, we have

$$L_n = \Phi(t_k, t_{k_n}) - \sum_{i=1}^{n-1} L_i \Phi(t_{k_i}, t_{k_n}). \tag{61}$$

The error covariance matrix of the fused track can be defined as

$$P_f(t_k|t_k) = E\{\tilde{X}_f(t_k|t_k)\tilde{X}_f(t_k|t_k)^{'}\} \tag{62}$$

which allows determination of the weights L_i which define the optimal asynchronous track fusion filter. Further details about this track fusion approach are found in [2] and [1]. The solution is encapsulated in theorems like this:

Theorem 1: The error covariance matrix of the fused track using the fusion rule is given by

$$P_f(k|k) = \sum_{i=1}^{n-1}\sum_{j=1}^{n-1} L_i M_{ij} L_j' + \sum_{i=1}^{n-1} L_i N_i + \sum_{i=1}^{n-1} N_i L_i^{'} + M_n. \tag{63}$$

Theorem 2 : The minimum mean square solution of the asynchronous track fusion problem using the fusion rule is given by

$$X_f(t_k|t_k) = \sum_{i=1}^{n} L_i X_i(t_{k_i}|t_{k_i}), \tag{64}$$

$$P_f(t_k|t_k) = LML^{'} + LN + N^{'}L^{'} + M_n, \tag{65}$$

where

$$[L_1\ L_2\ ...\ L_{n-1}] = -N^{'}M^{-1}, \tag{66}$$

$$L_n = \Phi(t_k, t_{k_n}) - \sum_{i=1}^{n-1} L_i \Phi(t_{k_i}, t_{k_n}). \tag{67}$$

The reason this general example has been presented is that it is indicative of all fusion algorithms; they are variations on a similar theme. With the proper redefinitions of matrices, both track and data fusion amount to the same thing: **weighted (positive semi-definite and the sum is normalized to one) combinations of data**. The results can be extremely complicated, difficult to understand, difficult to predict in terms

of performance, and difficult to determine the algorithm's underlying correctness, but they have the same underlying mathematical form when understood properly. Given this observation, an alternative suggests itself as a means of understanding these types algorithms based on maximum entropy analysis. An example is the Bar-Shalom-Campo Fusion Rule:

$$X_f(k|k) = X_2(k|k) + (P_{22} - P_{21})U^{-1}[X_1(k|k) - X_2(k|k)], \tag{68}$$

$$P_f(k|k) = P_{22} - (P_{22} - P_{21})U^{-1}[P_{22} - P_{12}]. \tag{69}$$

This result is simple enough to work out the equivalent interaction model for and will be the subject of future work.

IV. Concluding Comments

The translation: **Interaction Model** $\Leftrightarrow$ **Probability Model** provides a richer interpretation of "physics". This allows us to consider more general types of interactions that the *statistical mechanics crank* can be brought to bear. Fusion of weighted information from different "atomic composition" elements is equivalent to creating an interaction system with an underlying physics that we can strive to understand. There is always a design physics model based on interpreting the weights associated with the components of a composition. This provides an entirely new method to analyze Design mathematically from the perspective of inverse problems based on the MEP perspective. It also provides a new basis for the mathematization of design. This is an area that requires further development. The field has rich predictive capability, the question is how useful it is when applied to specific problem domains. Each field raises interesting problems, and even more interesting problems will be raised as the initial ones are solved.

References

[1] A. T. Alouani, J. E. Gray, and D. H. McCabe, "Theory of Distributed Estimation Using Multiple Asynchronous Sensors", T-AES Vol. 41, No. 2, April 2005.

[2] Y. Bar-Shalom and T. Fortmann, *Tracking and Data Association,* Academic Press, Inc., Orlando, FL (1988).

[3] Frank, O. and Strauss, D., "Markov Graphs", *Journal of the American Statistical Association*, Vol. 81, 832-842, 1986.

[4] Harary, F., *Graph Theory*, McGraw-Hill Book Company, 1972.

[5] Jaynes, E. T., *E. T. Jaynes: Papers on Probability, Statistics, and Statistical Physics*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1989.

[6] Jaynes, E. T., in *Maximum Entropy and Bayesian Methods*, edited by C.R. Smith, G. J. Erickson, and P.O. Neudorfer, Kluwer Academic, 1992.

[7] Jeffreys, Sir Harold and Jeffreys, Lady Bertha S., *Methods of Mathematical Physics, Third Edition*, Cambridge University Press, 1956.

[8] Kesavan, H. K. and Kapur, J. N., "*The Generalized Maximum Entropy Principle*", IEEE Trans. on Systems, Man, and Cybernetics, Vol. 19, No. 5, Sept/Oct, 1989.

[9] Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, Third Edition, McGraw-Hill Book Company, 1991.

[10] Park, J. and Newman, M. E. J., "*The Statistical Mechanics of Networks*", arXiv:cond-mat/0405566v1 25 May, 2004.

[11] Ross, S. M., *Introduction to Probability Models*, Fourth Edition, Academic Press, 1989.

[12] Shannon, C. E. and Weaver, W., *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, 1964. (Originally published in 1949, this contains a reprint of Shannon's 1948 papers that appeared in the *Bell System Technical Journal*.)

[13] Skvoretz, J., "*Complexity Theory and Models for Social Networks*", *Complexity*, Vol. 8, No. 1, 2003.

[14] Strauss, D., "*On a General Class of Models for Interaction*", *SIAM Review*, Vol. 28, No. 4, 1986.